

# Model Rasch untuk Penelitian Sosial Kuantitatif<sup>1</sup>

Oleh: Bambang Sumintono<sup>2</sup>

## Pendahuluan

Menurut Punch (2009) riset kuantitatif dalam ilmu sosial lebih dulu dikenal dibandingkan penelitian kualitatif. Pada masa 1800-an dimana kemajuan sains yang pesat menyebabkan ilmu-ilmu sosial tertarik untuk menerapkan kaidah yang digunakan dalam ilmu fisika, kimia dan biologi, yaitu dalam hal eksperimentasi dan pengukuran (*measurement*). Dalam aspek pertama dalam ilmu sosial dikenal disain penelitian berupa eksperimen, kuasi eksperimen dan survey (Punch, 2009; Creswell, 2012); sedangkan pada aspek kedua memunculkan teori tes klasik (*classical test theory* atau CTT) yang digagas oleh Spearman pada tahun 1904 (Alagumalai et al., 2005), serta pendefinisian pengukuran oleh Stevens pada 1946 (Mari et al., 2012) yang diterapkan secara luas pada berbagai ilmu-ilmu sosial.

Definisi pengukuran dari Stevens, yaitu “*measurement is the assignment of numerals to objects or events according to some rule*”, menjadi rujukan utama perkembangan psikometri setelah era perang dunia kedua. Definisi yang lebih bersifat umum menjadi rujukan dalam teknik pengukuran pada ilmu sosial seperti yang digunakan oleh Thurstone, Guttman dan yang paling populer adalah Likert. Teknik pengukuran adalah hal yang sentral dalam penelitian kuantitatif pada ilmu sosial, karena itulah titik tolak mendapatkan data berupa angka yang menjadi focus riset ini.

Terdapat berbagai kritik yang dialamatkan pada CTT. Schumaker (dalam Alagumalai et al., 2005: 11) menyebutkan bahwa koefisien reliabilitas (ukuran konsistensi pengukuran) CTT tergantung banyaknya sampel, skala pengukuran yang tidak linear, keterbatasan rentang skor serta keseimbangan korelasi negatif-positif. Bila dilihat definisi dari Stevens (Mari et al., 2012) pun jelas disebutkan bahwa jenis data yang didapat melalui teknik pengukuran yang menanyakan opini atau sikap, adalah nominal dan ordinal sehingga alat analisis yang bisa digunakan pun

---

<sup>1</sup> Makalah dipresentasikan dalam kuliah umum di Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, 21 November 2014

<sup>2</sup> dosen pada Institute of Educational Leadership, Universiti Malaya, Kuala Lumpur, Malaysia

email: [bambang@um.edu.my](mailto:bambang@um.edu.my) dan [deceng@gmail.com](mailto:deceng@gmail.com)

Blog: <http://deceng2.wordpress.com> (pengalaman dosen) dan <http://deceng3.wordpress.com> (rasch model)

terbatas. Bahkan operasi aritmatika dasar seperti tambah, kurang, kali dan bagi pun tidak bisa dilakukan karena angka yang didapatkan bukan lah bilangan bulat namun skor yang berupa data ordinal.

Kekurangan CTT kemudian diperbaiki dengan teori respon butir (*item response theory* atau IRT) dengan berbagai variasi parameter logistiknya (PL), salah satunya adalah 1PL yang dikembangkan menjadi model rasch. Tidak seperti CTT yang selalu bergantung pada skor, IRT tidak tergantung pada sampel soal/ Pernyataan tertentu dan abilitas orang yang terlibat dalam ujian /survey. Pada makalah ini akan dijelaskan secara singkat tentang model rasch, pengukuran yang objektif, serta aplikasi model rasch dalam penelitian kuantitatif dalam ilmu sosial.

### **Model Rasch**

Georg Rasch mengembangkan satu model analisis dari teori respon butir (atau *Item Response Theory*, IRT) pada tahun 1960-an biasa disebut 1PL (satu parameter logistic) (Olsen, 2003). Model matematika ini kemudian dipopulerkan oleh Ben Wright (Linacre, 2011). Dengan data mentah berupa data dikotomi (berbentuk benar dan salah) yang mengindikasikan kemampuan siswa, Rasch memformulasikan hal ini menjadi satu model yang menghubungkan antara siswa dan aitem (Sumintono & Widhiarso, 2013).

Sebagai ilustrasi, seorang siswa yang mampu mengerjakan 80% soal dengan benar tentu mempunyai abilitas yang lebih baik dari siswa lain yang hanya bisa mengerjakan 65% soal. Data tersebut (persentase) menunjukkan bahwa data mentah yang diperoleh tidak lain adalah jenis data ordinal yang menunjukkan peringkat dan tidak linier (Linacre, 1999; Mohd Saidfudin Masodi, tanpa tahun). Oleh karena data ordinal tidak mempunyai interval yang sama, maka data tersebut perlu diubah menjadi data rasio untuk keperluan analisis statistik. Sehingga bila seseorang mendapat skor 80%, maka nilai *odds ratio*-nya adalah 80:20, yang tidak lain adalah data rasio yang lebih tepat untuk tujuan pengukuran. Melalui data rasio ini Rasch mengembangkan model pengukuran yang menentukan hubungan antara tingkat kemampuan siswa (*person ability*) dan tingkat kesulitan aitem (*item difficulty*) dengan menggunakan fungsi logaritma untuk menghasilkan pengukuran dengan interval yang sama. Hasilnya adalah satuan baru yang disebut logit (*log odds unit*) yang menunjukkan abilitas siswa dan kesulitan aitem; sehingga nantinya dari nilai logit yg didapat, disimpulkan bahwa tingkat kesuksesan siswa dalam

mengerjakan soal sangat tergantung dari tingkat abilitasnya dan tingkat kesulitan soal (Olsen, 2003).

Untuk data yang berbentuk dikotomi, pemodelan Rasch menggabungkan suatu algoritma yang menyatakan hasil ekspektasi probabilistik dari aitem 'i' dan responden 'n', yang secara matematis dinyatakan sebagai (Bond dan Fox, 2007):

$$P_{ni}(x_{ni}=1/\beta_n, \delta_i) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$

dimana:  $P_{ni}(x_{ni}=1/\beta_n, \delta_i)$  adalah probabilitas dari responden n dalam aitem i untuk menghasilkan jawaban betul ( $x = 1$ ); dengan kemampuan responden,  $\beta_n$ , dan tingkat kesulitan aitem  $\delta_i$ .

Persamaan di atas dapat lebih disederhanakan dengan memasukkan fungsi logaritma dan menjadikannya:

$$\log (P_{ni} (X_{ni} = 1 / \beta_n, \delta_i)) = \beta_n - \delta_i$$

Sehingga probabilitas akan satu keberhasilan dapat dituliskan sebagai:

probabilitas untuk berhasil	=	kemampuan responden	-	tingkat kesulitan aitem
--------------------------------	---	------------------------	---	----------------------------

### Rasch Model dan Pengukuran Objektif

Dalam lingkup ilmu sosial, maka mendapatkan data berupa angka sumber yang bisa didapat biasanya dalam bentuk sikap (*attitude*) dan opini terhadap aitem pernyataan atau pertanyaan dalam satu instrumen yang diberikan. Instrumen tersebut dirancang dari variabel yang sudah didefinisikan secara memuaskan, kemudian diidentifikasi konstruk-konstruk yang relevan; dari sana lah aitem-aitem dibuat dan dikembangkan untuk bisa mengukur variabel yang dimaksud. Pada saat yang sama pilihan jawaban yang disediakan umumnya kemudian mengikuti pola penskoran yang dianut oleh teori test klasik (CTT). Dalam konteks model rasch, pola penskoran yang 'menetap' ini tidak lain adalah pengukuran yang hasilnya bergantung pada siapa yang diukur (*test dependent scoring*); sedangkan yang harus dilakukan dalam riset kuantitatif dalam ilmu sosial adalah pengukuran yang objektif (*objective measurement*).

Konsep pengukuran yang objektif dalam ilmu-ilmu sosial menurut Mok dan Wright (2004) harus mempunyai lima kriteria, yaitu:

1. Memberikan ukuran yang linear dengan interval yang sama;
2. Melakukan proses estimasi yang tepat;
3. Menemukan aitem yang tidak tepat (*misfits*) atau tidak umum (*outliers*);
4. Mengatasi data yang hilang;
5. Menghasilkan pengukuran yang *replicable* (independen dari parameter yang diteliti)

Dari kelima syarat tadi, sejauh ini hanya rasch model lah yang bisa memenuhi kelima syarat tersebut. Dengan kata lain kualitas pengukuran dalam ilmu sosial yang dilakukan dengan rasch model akan mempunyai kualitas yang sama seperti halnya pengukuran yang dilakukan dalam bidang fisika.

Bila dilihat lebih lanjut, skala logit (*log odds unit*) yang dihasilkan dalam model rasch adalah skala dengan interval yang sama dan bersifat linear yang berasal dari data ratio (*odds ratio*) dan bukannya data mentah skor yang didapat (1). Oleh karena itu proses estimasi abilitas seseorang ataupun tingkat kesulitan soal akan mempunyai nilai estimasi yang lebih tepat dan bisa saling dibandingkan karena mempunyai satuan yang sama (logit) (2). Berhubung algoritma yang digunakan akan melakukan pengurutan secara terstruktur antara responden dari abilitas tinggi ke rendah, yang secara bersamaan juga mengurutkan soal dari yang mudah ke yang sulit, maka adanya ketidaktepatan/konsistensi jawaban dari responden (*misfit*) ataupun pola yang diluar kebiasaan (*outlier*) akan mudah dideteksi; demikian juga untuk pola respon yang diterima satu soal tertentu (3). Pengurutan abilitas responden dan kesulitan soal secara terstruktur juga membuat model rasch dapat melakukan prediksi bila terdapat data yang hilang (4). Skala logit yang dihasilkan akan memunculkan nilai yang tergantung dari pola respon yang diberikan, bukannya pada skor awal yang ditentukan, sehingga rasch model akan selalu menghasilkan pengukuran yang independen (5).

Analisis dengan model Rasch menghasilkan analisis statistik kesesuaian (*fit statistics*) yang memberikan informasi pada peneliti apakah data yang didapatkan memang secara ideal menggambarkan bahwa orang yang mempunyai abilitas tinggi memberikan pola jawaban terhadap aitem sesuai dengan tingkat kesulitannya. Parameter yang digunakan adalah *infit* dan *outfit* dari kuadrat tengah (*mean square*) dan nilai terstandarkan (*standardized values*). Menurut Sumintono dan Widhiarso (2013), *infit* (*inlier sensitive* atau *information weighted fit*) adalah

kesensitifan pola respon terhadap aitem sasaran pada responden (*person*) atau sebaliknya; sedangkan *outfit* (*outlier sensitive fit*) mengukur kesensitifan pola respon terhadap aitem dengan tingkat kesulitan tertentu pada responden atau sebaliknya.

Riset kuantitatif dalam ilmu sosial selalu menghadapi kritik yang mendasar dalam hal pengujian instrumen risetnya. Uji kuantitatif instrument yang biasa dilakukan dalam CTT adalah indeks realibilitas (alpha Cronbach) yang hanya mengukur interaksi antara aitem dan person; bagaimana kualitas individual aitem tidak pernah bisa dilakukan karena tiadanya indeks pengukuran yang bisa dilakukan; saat yang sama untuk mendeteksi jawaban responden yang tidak konsisten pun tidak tersedia. Hal yang berbeda dengan teori test klasik, dalam rasch model analisis aitem dilakukan ke tingkat masing-masing aitem. Selain terhadap aitem, rasch model juga secara bersamaan menguji person (responden), dimana akan terlihat pola jawaban responden yang konsisten, yang cenderung untuk menyetujui (dalam instrument sikap) maupun mengidentifikasi jawaban yang asal saja (Sumintono & Widhiarso, 2013). Uji untuk instrument riset pun bisa dilakukan dalam bentuk uji dimensionalitas, skala peringkat yang digunakan maupun deteksi adanya bias dari aitem yang diujikan. Kesemua itu bisa dilakukan karena pada dasarnya model rasch memenuhi semua syarat pengukuran objektif.

## **Aplikasi Model Rasch dengan Winstep untuk pengujian Instrumen (Linacre, 2011)**

### ***1. Pengembangan Instrumen Pengukuran***

Pemodelan Rasch menjadi alternatif pengembangan instrumen pengukuran selain menggunakan teori klasik. Beberapa tahap yang biasanya dilalui dalam prosedur pengembangan instrumen pengukuran adalah:

- a) Verifikasi asumsi unidimensionalitas dan independensi lokal pengukuran
- b) Pengujian ketepatan butir-individu dengan model. Butir yang memiliki nilai ketepatan rendah dikeluarkan dari analisis. Analisis diulang lagi hingga semua butir memiliki ketepatan dengan model.
- c) Jika jumlah butir yang tersisa masih melebihi jumlah butir yang ditargetkan, maka kita dapat menyeleksi butir dengan berbagai pertimbangan, misalnya : (a) butir yang tidak overlap lokasinya dengan butir lain, (b) butir yang dapat meningkatkan reliabilitas pengukuran, butir yang opsi-opsi responsnya sesuai dengan urutannya (menelaah grafik karakteristik butir) atau (d) butir yang memberikan informasi yang sesuai dengan fungsi pengukuran (menelaah grafik fungsi informasi).

Proses evaluasi terhadap instrumen pengukuran merupakan proses analisis yang bersifat iteratif, yang dilakukan berulang-ulang hingga peneliti menemukan komposisi yang optimal, dimana semua kriteria dapat terpenuhi. Pada program Winsteps, unidimensionalitas terdapat pada fungsi **Item : dimensionality** dan ketepatan butir dengan model (*infit-outfit*) dan lokasinya (*measure*) dapat dilihat pada **Item: measure** maupun **Item: fit order**.

## **2. Deteksi Bias Pengukuran**

Butir maupun instrumen pengukuran dapat bersifat bias, yaitu ketika sebuah butir lebih memihak pada salah satu individu dengan karakteristik tertentu. Sementara itu individu dengan karakteristik oposisinya justru dirugikan. Misalnya, butir sebuah tes kecerdasan anak melibatkan gambar berupa salju untuk dikenali kejanggalannya. Bagi anak-anak yang pernah berinteraksi dengan salju, soal ini cukup mudah dipahami. Sebaliknya bagi anak-anak yang tidak berinteraksi dengan salju, soal ini sulit dipahami. Butir ini cenderung bias dalam mengukur, yang dalam psikometri dinamakan dengan butir yang terjangkit keberfungsian butir diferensial (DIF/*differential item functioning*). Pemodelan Rasch menyediakan menu untuk memfasilitasi peneliti yang hendak mendeteksi adanya butir-butir yang terjangkit DIF.

Pada paket program Winsteps, informasi mengenai bias butir ini dapat dilihat melalui **Item: DIF, between/within**. Butir-butir yang memiliki nilai P (PROB.) di bawah 0,05 menunjukkan bahwa butir tersebut terjangkit DIF pada tabel tersebut akan muncul nilai selisih tingkat kesulitan butir ditinjau dari dua sampel yang diuji. Misalnya jender, latar belakang budaya atau lokasi demografis; bahkan Winsteps juga bisa mendeteksi kombinasi data demografis misalnya jender dan lokasi demografis.

## **3. Deteksi Bias Individu**

Jika bias pengukuran disebabkan oleh performansi butir yang berbeda ketika diterapkan pada dua sampel berbeda (e.g. laki-laki & perempuan), maka bias individu disebabkan oleh performansi individu yang berbeda ketika berhadapan dengan butir yang berbeda. Misalnya peneliti memiliki dua tes yang sama, akan tetapi diadministrasikan dengan cara yang berbeda (paper-pensil vs. komputer). Idealnya diharapkan hasil yang didapatkan tidak berbeda, karena prosedur administrasi tes tidak terkait dengan atribut yang diukur. Ketika bias ini muncul, dimana ada beberapa individu yang memiliki skor tinggi hanya pada tes paper pensil dibanding dengan komputer, maka tes ini memiliki masalah dalam hal keberfungsian personal diferensial

(*differential person functioning/DPF*). Program Winsteps dapat mengakomodasi deteksi ini melalui **Person: DPF, between/within**. Individu yang memiliki performansi yang tidak konsisten pada butir yang diujikan akan dapat dideteksi melalui tabel tersebut. Beberapa penelitian juga menggunakan teknik ini untuk mendeteksi jawaban yang muncul dari kecurangan.

#### **4. Identifikasi Dimensionalitas Pengukuran**

Pengukuran adalah mengidentifikasi satu atribut yang dilakukan pada satu waktu. Untuk mengidentifikasi berapa atribut atau dimensi yang diukur, peneliti melakukan analisis faktor. Butir-butir yang memiliki korelasi tinggi akan mengumpul menjadi satu dimensi yang bisa jadi terpisah dengan butir-butir lain yang juga membentuk satu dimensi. Identifikasi dimensionalitas pengukuran berguna untuk mengoptimalkan pengukuran yang dilakukan sehingga informasi yang diberikan lebih memusat pada atribut yang diukur. Misalnya, mengukur intensitas pelanggan untuk membeli produk X, benar-benar menghasilkan informasi mengenai intensitas membeli, dan bukan kepuasan terhadap produk X.

Dalam analisis faktor, pemodelan Rasch memiliki perspektif yang berbeda dalam mengidentifikasi adanya dimensi–dimensi dalam pengukuran. Adanya dimensi tidak hanya dilihat berdasarkan keterkaitan antar butir, karena butir yang memiliki keterkaitan yang tinggi bisa jadi merupakan dimensi yang terpisah. Misalnya, tinggi badan dan berat badan memiliki korelasi yang tinggi akan tetapi keduanya merupakan atribut yang berbeda. Menurut perspektif pemodelan Rasch, suatu pengukuran dikatakan berdimensi majemuk ketika analisis menghasilkan korelasi antar residu (varian yang tidak dijelaskan oleh model) butir di atas 0,3. Residu diharapkan bersifat acak sehingga tidak berkaitan dengan residu lainnya. Jika korelasi antar residu yang dihasilkan adalah tinggi, maka ada atribut atau dimensi lain yang tidak dapat dijelaskan model.

Pada program Winsteps, dimensionalitas pengukuran dapat dilihat pada **Dimensionaliti maps**, yang menunjukkan berapa varians yang dijelaskan oleh model dan diluar model. Nilai *unexplnd variance in 1st contrast* yang dibawah 2 menunjukkan bahwa unidimensionalitas pengukuran telah dapat dibuktikan.

### 5. *Pengujian skala peringkat (rating scale)*

Analisis validitas skala peringkat adalah pengujian yang dilakukan untuk memverifikasi apakah peringkat (*rating*) pilihan yang digunakan membingungkan bagi responden atau tidak. Misalnya untuk teknik pengukuran yang menggunakan peringkat Likert (*Likert rating*) diberikan lima buah pilihan antara rentang STS, TS, N, S dan SS yang tidak lain adalah data politomi. Analisis model Rasch dapat memberikan proses verifikasi bagi asumsi peringkat yang diberikan dalam instrument yang digunakan. Pada program Winsteps, pengujian skala peringkat pengukuran menggunakan **Rating (partial credit) scale**. Hasil yang ditunjukkan adalah rata-rata observasi (*Observed Average*) dan *Andrich Threshold* yang menunjukkan ketepatan pilihan yang diberikan pada responden.

### Kesimpulan

Pengujian instrument dalam penelitian kuantitatif adalah hal yang esensial, berhubung dari sana lah data berasal, sehingga akan menentukan kualitas temuan riset. Model Rasch dapat banyak membantu peneliti ilmu sosial dalam meningkatkan kualitas riset yang dilakukan, karena prinsip dasar yang tepat dan model pengolahan data yang sesuai untuk riset kuantitatif khususnya dalam pengolahan data ordinal. Hal ini karena model Rasch sesuai dengan lima persyaratan pengukuran yang objektif.

### Daftar Pustaka

- Alagumalai, S., Curtis, D.D. and Hungi, N. (editors) (2005). *Applied Rasch Measurement: book of exemplars*. papers in honour of John P. Keeves. Dordrecht: Springer.
- Bond, T.G., & Fox, C. (2007). *Applying the Rasch Model. Fundamental measurement in the Human Sciences*. Lawrence Erlbaum Associates, Publishers. Mahwah. New Jersey
- Creswell, J.W. (2012). *Educational Research: planning, conducting and evaluating quantitative and qualitative research*. Boston: Pearson.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*. 3(2), 103-122.
- Linacre, J.M. (2011). *A User's guide to WINSTEPS Ministeps*; Rasch-model Computer Program. Program Manual 3.73.



- Mari, L., Carbone, P. and Petri, D. (2012). Measurement fundamentals: A pragmatic view. *IEEE Transactions on Instrumentation and Measurement*, 61(8), 2107-2114.
- Mohd Saidfudin Masodi. (tanpa tahun). Instrument Construct Validation of VFMA\_35 for the conduct of Value for Money Audit using Rasch Model. Makalah tidak diterbitkan, Accounting Research Institute, UiTM, Syah Alam, Malasia.
- Mok, M. and Wright, B. (2004). Overview of Rasch Model Families. **In** *Introduction to Rasch Measurement: Theory, Models and Applications* (hal 1-24). Minnesota: Jam Press.
- Olsen, L. W. (2003). Essays on Georg Rasch and his contributions to statistics. Unpublished PhD thesis at Institute Of Economics University of Copenhagen.
- Punch, K. F. (2009). *Introduction to Research Methods in Education*. Sage: London.
- Sumintono, B dan Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: Trim Komunikata Publishing House.